KTH Computer Science
and Communication

Simulating the evolution of silverfish

Evolution modelled as an evolutionary algorithm

ERIK NORELL eriknore@csc.kth.se
DANIEL CHAVEZ dace@csc.kth.se

Royal Institute of Technology
DD143X, Bachelor's Thesis in Computer Science
Supervisor: Vahid Mosavat
Examiners: Karl Meinke and Örjan Ekeberg

29 april 2014

Abstract

An area of Artificial Intelligence, used for instance in optimization, is evolutionary algorithms. By using mechanisms similar to those that cause evolution, evolutionary algorithms can improve e.g. problem solving algorithms by artificial evolution. The purpose of this study was to show that it's possible to simulate evolution by modelling it as an evolutionary algorithm. This was achieved by simulating the evolution of silverfish' genes in two environments with the only difference of the presence of a threat. The results were considered to be successful as the majority of the genes which were presumed to be important for survival changed in such a way. The results could be repeated between simulations indicating that random change of the genes and deterministic factors in the environment shaped the genes of the silverfish and that after simulation the silverfish were optimally fit for the environment.

Sammanfattning

Ett område inom Artificiell Intelligens som bl a används för optimering är evolutionära algoritmer. Genom att använda mekanismer liknande de som orsakar evolution så kan t ex algoritmer förbättras genom artificiell evolution. Syftet med den här studien var att visa att det är möjligt att simulera evolutionen genom att modellera den som en evolutionär algoritm. Detta gjordes genom att simulera evolutionen av silverfiskars gener i två miljöer vars enda skillnad var en förekomst av ett hot. Resultaten ansågs vara lyckade då merparten av de gener som förmodades vara viktiga för överlevnad förändrades på ett sådant sätt. Resultaten kunde upprepas mellan simuleringar vilket indikerar att genom stokastisk förändring av gener och deterministiska faktorer i miljön så optimerades silverfiskarnas gener för att göra det möjligt att överleva i respektive miljö.

Contents

1 Introduction

Artificial intelligence (AI) is created intelligence exhibited by machines and computer software. It is a common area of research at universities and companies around the world. A sub-branch of AI is Computational Intelligence which is mostly concerned with optimizing problems which are impossible or computationally too resource consuming to solve exactly [1]. One way of optimizing is Evolutionary computation where optimization is done using evolutionary algorithms. Evolutionary algorithms is a flexible and adaptable way of finding better solutions to problems. It has a wide range of applications in areas of engineering design (e.g. aircraft design, electromagnetic systems, nuclear fuel arrangement optimization) and traditional difficult types of problems in computer science (e.g. travelling salesperson problem, knapsack problems, set partitioning and scheduling).

Optimizing a solution using evolutionary algorithms is a process of "artificial selection", mimicking the process of natural selection [3]. Starting with a population of problem solving algorithms (candidates) relevant to the problem, new candidates can be generated by combining elements of two candidates or changing a small part of a candidate. Then evaluating all candidates according to a best solution and repeating the process until a time limit has been exceeded or no improvements are found. This is similar to natural selection where in nature there are inheritance of and mutations on DNA and the principle of "survival of the fittest".

Problem Statement

The main purpose of this study is to investigate if it's possible to simulate evolution based on the main principles of evolutionary algorithms. More importantly, will silverfish be optimized to their living environment after the simulation?

For this study a simplified computer simulation of silverfish and evolution modelled as an evolutionary algorithm will be created. If successful, it would show that survival of the fittest from evolution is also a solution to an optimization problem. Successful results implies support to using evolutionary algorithms for optimization using variants of artificial selection but also support for the theory of evolution itself. The ambition is to create a genetically diverse population of silverfish and setting up a deterministic environment. The mechanisms behind the genetic variation in new generations of silverfish will be stochastic while components of the environment act deterministically on this variation, thereby optimizing the silverfish' ability to survive and reproduce under these conditions mimicking real evolution.

2 Background

In this section, the mechanisms behind evolution are described as well as central components of evolutionary algorithms and it's relation to evolution and natural selection. Also different types of simulation are described and compared.

2.1 Evolution and natural selection

As mentioned in the introduction evolutionary algorithms are based on natural selection and the theory of evolution. An individual's phenotype is the description of behavioral and physical characteristics, e.g. color, size, behavioral patterns. An individual's genotype is the actual physical DNA the individual has inherited [4]. It is a complex combination of these factors which decide an individual's fitness regarding survival.

Note that natural selection only occurs given an environmental pressure [2]. If resources needed for survival and reproduction are limitless and no external threats exists, all individuals should live long and be able to reproduce. The population would increase indefinitely. But given a limit on e.g. food, those individuals, or perhaps rather generations, with a geno- and phenotype resulting in a higher probability of finding food will also have a higher probability of survival. In other words, these individuals or generations are more fit given the environmental pressure. Thereby individuals with this particular geno- and phenotype have a higher probability of living long and have more offspring than individuals with less favorable geno- and phenotypes. Thereby the relative occurrence of the more favorable geno- and phenotype should increase over time. This is what is called natural selection, i.e. the survival of the fittest [2].

2.2 Evolutionary algorithms

Figure 1 shows a limited overview of AI and some subfields including those involving evolutionary algorithms. The research area of Evolutionary computation consists of four subareas: Genetic Algorithms, Evolution strategies, Evolutionary programming and Genetic Programming [1]. The first three subfields emerged independently of each other during the sixties and seventies. Not until the early nineties were they considered the same area of research. About at the same time the fourth subarea emerged. The differences between these areas concerns mainly area of application and/or implementation of evolutionary algorithms. For the purpose of this study, there is no need to differentiate these areas further. Focus lies on the area of evolutionary algorithms.

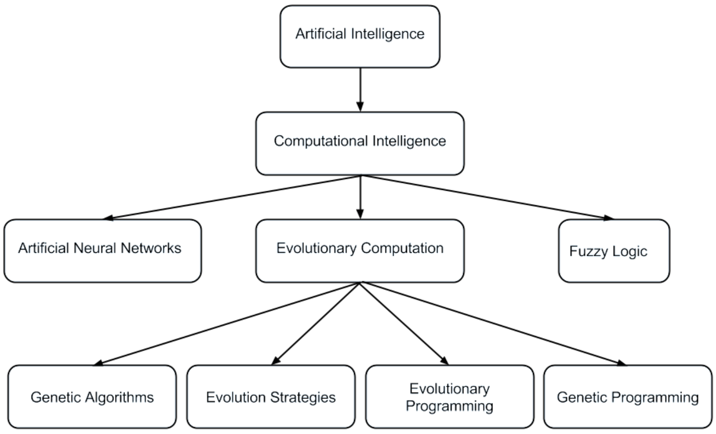


Figure 1: *Table over Computer Intelligence, a sub branch of Artificial Intelligence, and relevant areas of research.*

Fundamental when using evolutionary algorithms is modeling the problem in a way so that candidate solutions can be altered. It is also essential to have some form of evaluation of the different generations of candidates, most often some kind of function or selection-parameters are used to rank the candidates according to how close to a solution they are [3].

There are two methods used to generate new candidates [3]. Crossover means combining parts of two or more different candidates to form a new candidate. The other method is mutation, i.e. parts of a candidate solution is stochastically altered to generate a new candidate. How and to what extent these methods of regeneration are possible is dependent on how the problem and the candidate solutions are modeled [1].

Some extra care has to be made regarding the variation of candidates. When initializing the evolutionary algorithm one can inoculate certain candidates known in advance to be good, so as to assure they are not prematurely excluded [3]. There is also often a need for some extra mechanism to preserve a diversity of candidates when applying the evaluating function, otherwise generations similar to the best candidate with small variations might replace the other candidates within a relatively short period of time [3]. This can be achieved with a selection process where candidates are selected with a probability related to its fitness, or to randomly select groups of candidates and then select the most fit candidate from each group. Finally there must be some kind of termination criteria, most often a time threshold is used or when no better candidate within a timeframe has been found [3].

With all these parameters there is often a need for calibration of the algorithm. This can be done by time-consuming trials, using statistical analysis or automatically while solving the problem [3].

2.3 Simulation

There are several methods available when one wishes to study the behavior of a real-life system or predict a future state of the system. These methods can be mathematical analysis on the system, analyzing a prototype of the system, experimenting on the actual system or simulating the system with a computer [7]. The latter one, computer simulation, is applicable to many systems and is widely used [6].

Several limitations and problems arise with the alternative methods, mathematically analyzing a complex system is often impossible and experimenting with prototypes or the system can be both costly and time consuming [7][5]. Real-life systems often involve many parameters and functions which involve the element of randomness, this is where computer simulation is applicable [7][6].

Computer simulation is classified in to two designs: continuous event simulation and discrete event simulation. Either designs can be used in a simulation depending on on what aspect of the system is being studied.

Continuous event simulation is appropriate when parameters change continuously because the system's events are continuous. For example, the temperature T in a room at the time t. These systems are usually represented by differential equations which can not be solved analytically and the simulation involves finite-difference equations [7].

In discrete event simulation, the system's events occur in a discrete manner which means that the system's state and parameters are discrete. For example, the amount of people P in a room at the time t. These systems can be seen as the mathematical step function and are usually simulated with a sequential algorithm [5]. Discrete event simulation is the most common simulation design [6], however it does have limitations due to its sequential nature, discrete event simulations require concurrent approaches when the amount of events being simulated are high [5].

3 Initial considerations and hypotheses

The simulation is built according to an evolutionary algorithm, with silverfish as candidates, the environment as the evaluative function and with the goal of finding an optimal genetic variation of silverfish to survive in the given environment. As simulated evolution is based upon evolutionary algorithms, it can be regarded as an evolutionary algorithm by itself.

3.1 Methodological considerations

Two methodological issues must be addressed in order for this study to fulfill it's purpose:

1. How can it be determined if the silverfish are adapting to the environment and that change is simply not happening at random?

2. How can it be determined that a change is optimal?

One way to resolve the first issue is to do measurements before and after the simulation to show that a change has indeed occurred. But also run the simulation under two different conditions and that each condition presumably affects only some gene or genes. To address the second issue a number of simulations are run and compared. Given that the same environment is used for each simulation and that the environment acts deterministically on the silverfish' genes, repeated simulations should lead to similar results if the genes are indeed optimized.

3.2 Implementational considerations

Two types of environments will be simulated and hopefully show that the differences in environment determines the occurrence of genes which are necessary for survival. In the first environment there will not be any threats to the silverfish other than risk of starvation. In the other environment there will also be a predatory threat, from time to time a human comes with a flashlight killing all silverfish he or she sees.

Each silverfish will have a set of genes. The genes will determine the behavior of the silverfish, the ability to reproduce. and ultimately survival. The environment will consist of a limited geographical area with varied access to food. For the first scenario, there will not be an external threat, the genes important for survival should be the ones affecting the ability to find food, the lifespan and reproductive ability. For the second scenario, a physical threat is introduced, and genes important for avoiding the threat should be more favorable.

3.3 Hypotheses

The set of genes for each silverfish will be limited to three categories

1. Genes linked with general survival, i.e finding food, ability to reproduce, lifespan, etc.
2. Genes linked with ability to detect and avoid threats
3. Genes not linked to survival

Simulating two different environments, the expected result is that without the threat of the second scenario, only genes belonging to category 1 will be important to survival. In other words, after simulation the average silverfish should score high on genes in category 1. With "score high" meaning genes with high ability to find food, reproduce, etc. *With* the threat, the expected result is that genes from both category 1 and 2 are important, i.e. the average silverfish should score high on genes belonging to category 1 and 2. More precisely the expected result is that only genes belonging to category 2 should differ between the environments. Genes belonging to the third category should vary randomly between all simulations as these shouldn't be important for survival. The last category of genes can be seen as a measure of control that randomness exists within the simulation. To summarize the hypotheses:

- for the simulation *without* the threat: genes from category 1 will be optimized while genes in category 2 and 3 are unaffected by evolution.

- for the simulation *with* the threat: genes from category 1 *and 2* will be optimized while genes in category 3 are unaffected by evolution.

As already mentioned, several simulations should result in similar variations between each simulation. By making a number of simulations for each scenario, we hope to show by means of statistical analysis that our hypotheses are correct and that these results are consistent.

4 Method

The following sections describe how the silverfish and the environment was modelled in relation to evolution or perhaps rather natural selection. It's a description of the type of simulation and how it is constituted. Also how environmental factors and genes affect the silverfish behavior.

Since no similar studies or simulations have been found, the model is based on our assessment of what is reasonable and practical given the object and timeframe for this study. The intention is not to simulate actual silverfish, these can be regarded simply as living beings in an environment that limits their ability to survive. It could have been any kind of being. Therefore, the modelling of silverfish will not necessarily reflect real silverfish. The silverfish genetic makeup has been chosen according to the complexity of the simulation and in a way so that statistical calculations on the genetic variation of the population is possible. Genotypes and phenotypes are not distinguished as it would become too complex given the object of the study. We let a gene directly affect a behavior or form a characteristic, e.g. lifespan or energy level.

An example of how this simulation is different from reality is how and to what extent silverfish reproduce. In the simulation when two sexually mature silverfish mate, they only get one offspring rather than several. It is how *often* the silverfish can reproduce that determines the amount of offspring each silverfish generate, this in turn is linked with when the silverfish becomes sexually mature, the level of energy required to mate, lifespan and so on. In reality silverfish lay eggs and do so only in humid and relatively warm environments.

4.1 Simulation

The type of simulation that fit the object of this study best is a discrete event simulation, since a continuous approach is beyond the timeframe of this study, neither are we interested of stopping in the middle of a running simulation. For all silverfish, their respective alternative courses of action are evaluated and performed (movement, reproduction, food consumption or death). Thereafter the birth of new silverfish and regeneration of food are performed. Lastly the human's movement is generated. All these actions correspond to one discrete event. A simulation ran for 100 000 turns, approximately corresponding to around 40-50 generations of silverfish.

In the simulation the first environmental pressure is the limited availability of food. Silverfish with favorable genotype for finding food and reproducing will do so at the expense of other silverfish that have not as favorable genotype. This corresponds to the evaluative function for selection of candidates in evolutionary algorithms. It's similar to what happens during natural selection and evolution, although in reality it is of course more complex.

For the second scenario both the limited resource of food and the threat of the human acts as the environmental pressure. The human appeared a fixed number of times and for a fixed length of turns. The human appearance was represented as a flashlight and every silverfish that happens to be in the light dies immediately. Silverfish with a genotype enabling it to detect the human from a longer distance should have a better chance to survive and thereby reproduce, at the expense of silverfish with a lesser ability.

As mentioned earlier, in evolutionary algorithms there is a need to take precautions so that alternative solutions are not discarded prematurely, i.e. a certain variety of candidates is preserved. Since our evaluative function is more or less passive and continuous, we don't not need to take these precautions. It is rather a calibration problem which is discussed later on.

As explained in more detail under *Silverfish*, each silverfish has genes that decide traits. These genes vary in value between 0 and 1. At the beginning of each simulation the value for each silverfish and each gene was sampled from a normal distribution with a mean of 0.5 and a standard deviation of 0.125. The advantage of this is a good mix of values while avoiding too extreme values in the beginning. If we had chosen a uniform distribution of the values of the genes, many more silverfish would die early on in the simulation. By choosing a mean of 0.5 the values can increase or decrease during evolution.

4.1.1 Environment

For the environment a map in the form of a matrix with the x and y coordinates (called tiles) was used. In each tile there could be an unlimited number of silverfish or one unit of food.

Area: 300×300 pixels or tiles
Population: 100 individuals
Food: initially 200 units and maximum around 600 units

The availability of food was regulated by regenerating 20 units of food every 10 turns as long as two conditions were fulfilled. The first was that the total amount of food didn't exceed three times the starting amount of food, which equals 600 units. The second condition was that the total amount of silverfish didn't exceed the starting amount, which equals 100 silverfish. Note that a maximum number of individuals was not set, rather this is indirectly regulated by the available amount of food. Each unit of food would increase the silverfish's energy level with 30 units. Why these figures and conditions were chosen is explained further on under *Calibration*.

It should be noted that this model is not entirely geometric as silverfish and the human can move diagonally at the same cost as moving straight. The length of the diagonal between two points is longer but for ease of programming it is regarded as the same distance as between two points straight from each other. This should not affect the simulation.

4.1.2 Silverfish

Each silverfish had a set of traits:

1. *Lifespan* (100-3000 turns)
2. *Energy level* (0-2500 units)
3. *Perception* (1-60 tiles)
4. *Energy level needed for reproduction* (0-600 turns)
5. *Reproductive cycle* (100-750 turns)
6. *Sensitivity to light* (10-200 tiles)
7. *Color* (0-1 on a grayscale)
8. *Sex* (Female or Male)

Within parentheses is the possible interval that each trait varies within, for sex it's either male or female. Regarding energy, one unit corresponds to one turn, i.e. it will decrease with one unit per turn. When the silverfish reaches it's maximum limit of 2500 units, it will not look for food until the energy level drops below 2400 units. Perception concerns the ability to detect food and other silverfish of the opposite sex in the surroundings of the silverfish. Reproductive cycle refers to the minimum amount of turns between breeding, this amount of turns also works as age of sexual maturity. A silverfish is able to reproduce if it has enough energy and is fertile. Note that speed was not a trait but the same for all silverfish, 1 tile per turn.

Each trait is decided by the silverfish correspondent gene, except sex which was decided stochastically with an equal chance of being male or female. The correspondent genes are:

1. *Lifespan*
2. *Starting energy level*
3. *Perception*
4. *Energy level needed for reproduction*
5. *Reproductive cycle*
6. *Sensitivity to light*
7. *Color*

Note that the gene relating to energy level decides the energy level that the silverfish starts with. All silverfish share a maximum energy level of 2500 units, while the minimum and maximum level to start with was 100 and 600 respectively depending on the value of the gene. All genes can have a value between 0-1 in steps of 10^{-16}, i.e. 16 decimals.

New silverfish are spawned when two silverfish, of the opposite sex and with the ability to reproduce, meet at the same point in the map. Reproducing takes 50 turns, i.e. the silverfish reproducing will be static for 50 turns. The offspring inherit, so to speak, their parents' genotype by taking the average of each of the parents' genes added together (crossover from evolutionary algorithms). Afterwards it is stochastically determined if a mutation occurs, which genes it affects and to what extent (mutation from evolutionary algorithms). The chance of a mutation was set to 10%. There was an equal chance of mutation affecting only one gene to all of the genes, i.e. 1 in 7 that one gene was affected, 1 in 7 that two genes were affected, and so on. Lastly the maximum amount of mutation was set to ± 0.30 but a gene could never have a value more than 1 or less than 0. As mentioned, which genes affected and to what extent was also decided stochastically.

Rules for course of action

A silverfish's behavior is determined by rules for what is possible, i.e. deciding factors are the silverfish's energy level, what's physically around the silverfish, and what the silverfish can perceive, specifically food, other silverfish or the human's flashlight. The possible courses of action are given different priority.

Two silverfish that are reproducing are not able to move. Otherwise the most prioritized course of action is avoiding the human. Given that the silverfish can detect the human's flashlight, it will move in the opposite direction of where the human is. The second most prioritized course of action is reproduction. If the silverfish is able to reproduce and can perceive another silverfish in the same situation, then the silverfish will move towards each other and eventually reproduce. If conditions for none of these situations are fulfilled the silverfish will search for food. If the silverfish can perceive food in it's vicinity, it will move towards the food. Otherwise it will move towards a randomly set point on the map until one of the aforementioned courses of action is possible. Note that if a silverfish detects the human flashlight it will run away regardless of energy level and so on.

4.1.3 Human

The number of times that the human with a flashlight appeared as well as the length of each occurrence was fixed and set before the simulation began. The human appeared every 10% of the total length of the simulation, i.e. appeared every 10 000 turns. The length of the appearance was that same interval divided by 5, i.e. the length of each appearance was 2000 turns. In total that means that the human was present 18% of the time.

As already mentioned, the human was represented or modelled as a flashlight, any silverfish within the light died immediately. The size of this flashlight was 30×30 tiles and square. The shape of the flashlight shouldn't matter and from a programming view it was less trouble with a square than a round flashlight. For the flashlight to be effective at killing silverfish the speed of the flashlight was 3 tiles per turn to compare with the speed of the silverfish 1 tile per turn.

The human's movement was completely random, a random target on the map was set and when the human reached this target a new target was generated.

4.1.4 Specifying the hypotheses

We have divided the genes into three categories depending on how we expect the genes to change during evolution.

Category 1 linked to the ability to find food and reproduce

1. Lifespan (higher)
2. Starting energy level (higher)
3. Perception (higher)
4. Minimum energy level needed for reproducing (lower)
5. Reproductive cycle (lower)

In parentheses is how we predict the values of the gene will change during evolution. How these genes are linked to the ability to survive and to reproduce is in most cases more or less evident, e.g. a longer life implies a greater chance to reproduce more often and hence a greater spread of higher value of the genotype.

Category 2 linked to the ability to detect and react to threats

6. Sensitivity to light (higher with human, unchanged without human)

The idea is that the higher the sensitivity to light, the greater caution the silverfish will show, i.e. a tendency not to move in that direction which should enable the silverfish to live longer and be able to reproduce. This behavior is linked to whether the human is present or not. So a difference is expected on this gene depending on whether the simulation is run with the human present or not. For simulations without the human this gene shouldn't be affected by evolution, i.e. the expected mean is as in the beginning around 0.5 with a similar distribution, a standard deviation around 0.125.

Category 3 genes that should not affect the silverfish's survival

7. Color intensity (unchanged)

This gene corresponds to the silverfish color in grayscale, 0 being white and 1 being black. This gene does in no way affect the silverfish's behavior or ability to survive. This gene can be seen as a control-variable that there is enough randomness within the simulation and that increase or decrease in other genes are not by chance. The expected mean after evolution is close to it's initial value and standard deviation, i.e. 0.5 and 0.125 respectively.

4.2 Material

The programming language used for writing the simulation was Java. The statistical analysis was made using GNU PSPP[1] which is a free alternative to IBM's software with the same purpose SPSS. The figures were made using Matlab[2]. As this report is written, the code used for the simulation is available from Github[3].

During development, a visual representation of the simulation was used for debugging purposes shown in Figure 2. It was of great help to see that the silverfish and human behaved as expected. It's also a practical demonstration tool.

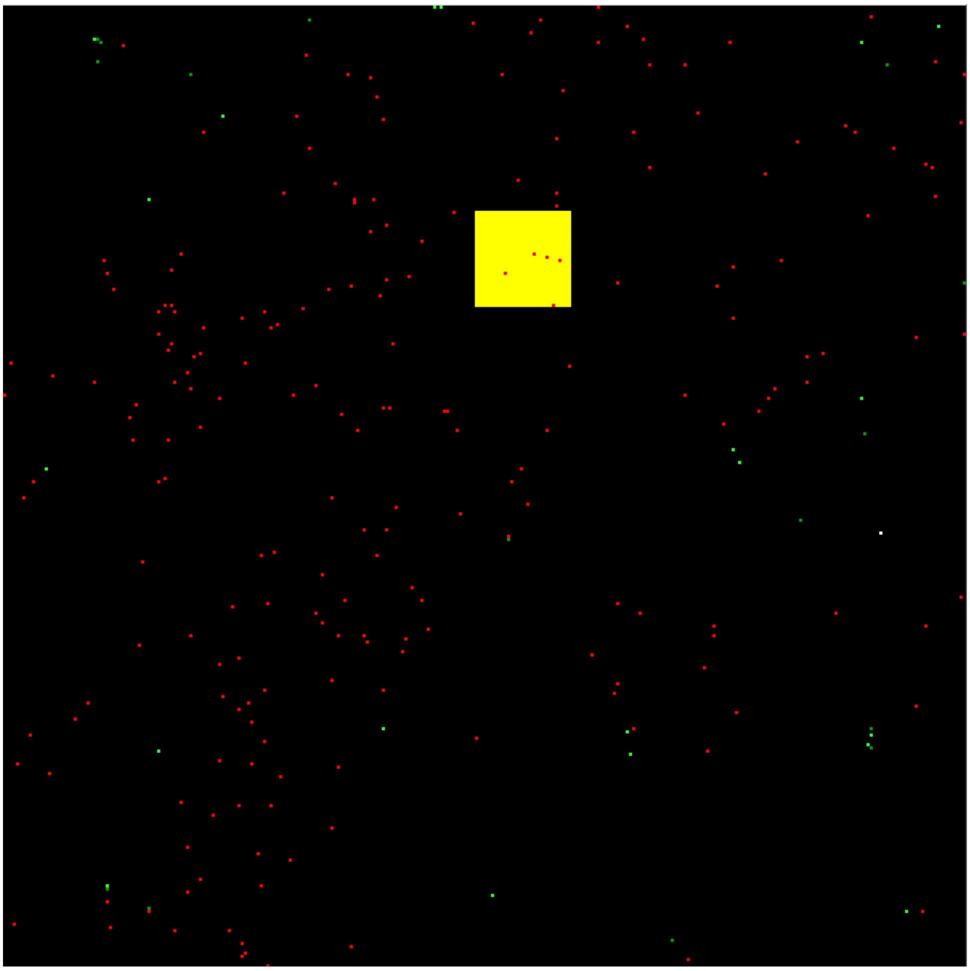


Figure 2: *Screenshot of the visualization of the simulation. The yellow square is the human flashlight. Red dots are food and green dots silverfish. Females have a darker green color than males. White dots indicate silverfish reproducing.*

[1] https://www.gnu.org/software/pspp/
[2] http://www.mathworks.com/products/matlab
[3] https://github.com/dacechavez/silverfi

4.3 Calibration

The maximum and minimum values for the traits were set to what seemed reasonable and were *not* an object for calibration, although some changes to lifespan and starting energy were made to enable a stable simulation. When calibrating, different sizes of the map were tested, different amount of silverfish to start the simulation, different amount and effect of food and, regarding the human, the size and the speed of the flashlight.

To get reasonable results we wanted the amount of silverfish before and after to be similar. We also wanted a fairly small runtime on the simulations as we intended to run multiple simulations. We found that a map with a size of 300×300 tiles and around 100 silverfish was good for a fast simulation. Running multiple simulations and combining the populations would still give us a large enough population to draw conclusions from. However we did have problems controlling the size of the population. At first we had a fixed interval of regeneration of food and the amount of food regenerated. Changing these values as well as how much increase in energy a unit of food gave a silverfish, only resulted in all silverfish dying early in the simulation or the population increasing uncontrollably. Therefore we set the two conditions as mentioned earlier which kept the population under control, still with the limitation of food as the environmental pressure.

When running simulations with the human present we wanted the human to have an impact on the population. If he or she didn't have an impact then the human would have no impact on the evolution of the silverfish either. Too much impact and all the silverfish would be exterminated. By trial and error we found a reasonable size and speed for the flashlight as well as a reasonable number and length of appearances.

5 Results and analysis

The primary results concerns the evolutionary changes of the silverfish' genes, as well as the comparisons between simulations. The secondary results are primarily to ensure that the simulation was successful, containing general results from the simulation.

5.1 Primary results

A total of 30 simulations were conducted with each simulation lasting 100 000 turns. Each simulation had a starting population of 100 silverfish and with the given constraints the ending populations were close to 100 silverfish.

5.1.1 Evolution of genes

Figure 3 and Figure 4 shows typical distributions for a gene that *has* changed during the simulation and a gene that has *barely* changed during simulation respectively. They are included for illustrative purposes and to provide the reader with a better understanding of the tables presented later on. Mean differences are reflected by a horizontal shift while differences in standard deviation is reflected by a difference in amplitude and horizontal distribution.

Figure 3 shows the distribution of values for the gene concerning starting energy before and after a simulation. The red normal distribution curve is based on the mean and standard deviation for the values of the gene.

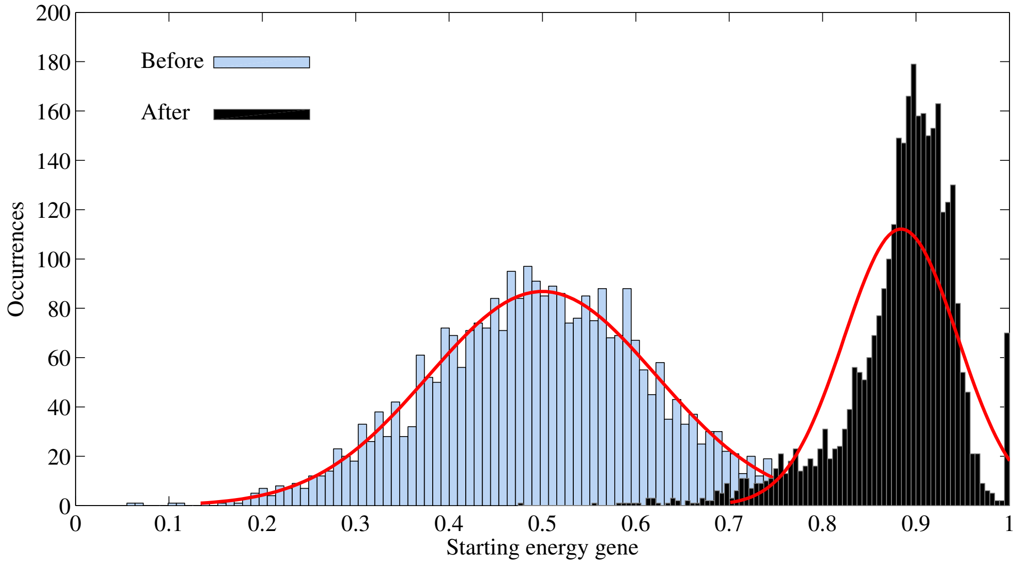


Figure 3: *Distribution of values for the gene concerning starting energy.*
Before: $mean : 0.50$; $std.dev : 0.12$.
After: $mean : 0.88$; $std.dev : 0.06$.

Figure 4 shows the values corresponding to Figure 3 for the gene concerning perception. Notice that the mean is roughly the same and the standard deviation decreased only slightly.

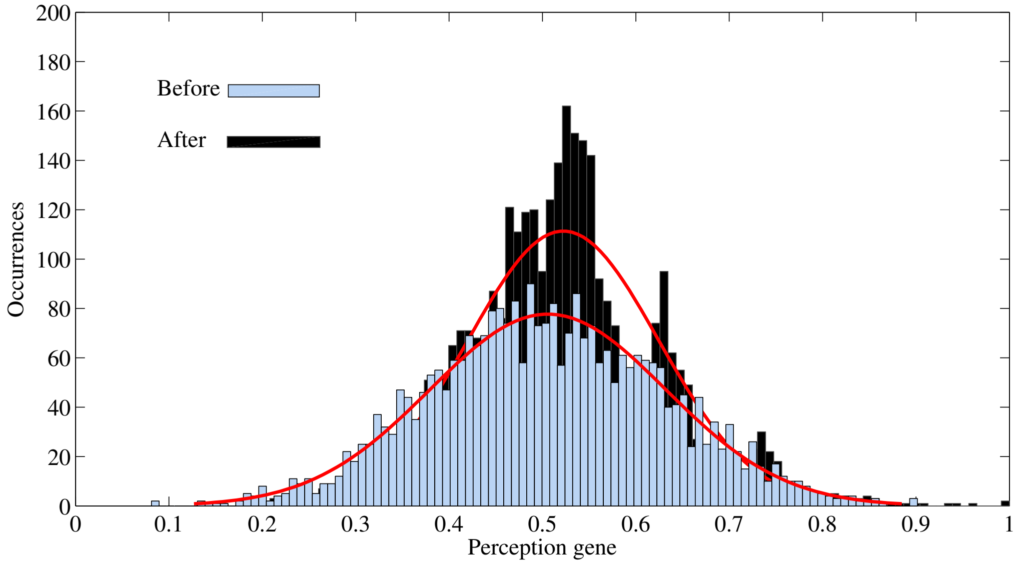


Figure 4: *Distribution of values for the gene concerning perception.*
Before: $mean : 0.50$; $std.dev : 0.12$.
After: $mean : 0.52$; $std.dev : 0.10$.

In Table 1 below, *Before* and *After* denotes the first and last turn respectively. As the amount of silverfish was fixed before simulation the total amount of silverfish was 3000, while after simulations this figure varied, hence the total amount of silverfishes was 3256 after the simulations.

Looking at the last column, the mean differences, there was a notable change for genes 1, 2, 4 and 5. A slight change for gene 6, but notice that for these figures both simulations *with* and *without* a human are pooled together. There is barely any change for gene 3 and 7. Also notice that most often the standard deviation has decreased for the genes where the mean has changed. All differences are significant but as the groups for the t-tests (N) are rather large, these figures are not that conclusive as even very small differences are significant (e.g. Color in Table 1).

Gene	**Time**	**N**	**Mean**	**Std.dev**	**Mean difference**
1. Lifespan	Before	3000	0.50	0.13	-0.23**
	After	3256	0.73	0.11	
2. Starting energy	Before	3000	0.50	0.12	-0.38**
	After	3256	0.88	0.06	
3. Perception	Before	3000	0.51	0.13	-0.02**
	After	3256	0.52	0.10	
4. Reproduction	Before	3000	0.50	0.12	0.38**
	After	3256	0.12	0.10	
5. Rep. cycle	Before	3000	0.50	0.12	0.43**
	After	3256	0.07	0.07	
6. Light sensitivity	Before	3000	0.50	0.13	0.09**
	After	3256	0.41	0.14	
7. Color	Before	3000	0.50	0.12	0.01*
	After	3256	0.50	0.12	

Table 1: *T-tests comparing the values of genes before and after simulations. The results include all populations from 30 simulations.*

* indicates $p < 0.05$ and ** $p < 0.01$

As mentioned, PSPP was used for the statistical analysis. All results are rounded to two decimals automatically by PSPP. However this rounding sometimes doesn't add up, e.g. the gene color: the means before and after are both 0.50 but the difference is still shown as 0.01. As we could not get all decimals we present the same figures as presented by PSPP.

Out of the 30 simulations, 15 were conducted *with* a human and 15 *without* a human. In Table 2 below, the genes after the simulations and for simulations *with* and *without* the human are compared, in order to see if the human had any effect. As can be seen, the only notable change was for gene 6.

Gene	Human	N	Mean	Std.dev	Mean difference
1. Lifespan	No	1631	0.74	0.11	0.02**
	Yes	1625	0.72	0.12	
2. Starting energy	No	1631	0.89	0.06	0.02**
	Yes	1625	0.88	0.06	
3. Perception	No	1631	0.51	0.08	-0.02**
	Yes	1625	0.53	0.12	
4. Reproduction	No	1631	0.12	0.11	-0.01**
	Yes	1625	0.13	0.09	
5. Rep. cycle	No	1631	0.07	0.07	-0.01**
	Yes	1625	0.08	0.06	
6. Light sensitivity	No	1631	0.45	0.18	0.09**
	Yes	1625	0.36	0.08	
7. Color	No	1631	0.49	0.10	0.00
	Yes	1625	0.50	0.14	

Table 2: *T-tests comparing end of simulation results between 15 simulations with human and 15 simulations without human.*

5.1.2 Between simulations

For every simulation a mean for each gene was calculated. These means represents the average silverfish for each particular simulation. Comparing these shows how large variations there are between simulations. It also serves as a better indication of if a gene has changed or not. In Table 3 below, these figures are presented for *before* simulation and *after*.

As can be seen in the last column of Table 3, differences between before and after simulation was observed for gene 1, 2, 4, 5 and 6. Again notice that simulations *with* and *without* the human are pooled together regarding gene 6. No differences was observed for gene 3 and 7.

Gene	Time	N	Mean	Std.dev	Mean difference
1. Lifespan	Before	30	0.50	0.02	-0.24**
	After	30	0.74	0.09	
2. Starting energy	Before	30	0.50	0.01	-0.38**
	After	30	0.88	0.05	
3. Perception	Before	30	0.51	0.01	-0.01
	After	30	0.51	0.08	
4. Reproduction	Before	30	0.50	0.01	0.37**
	After	30	0.13	0.11	
5. Rep. cycle	Before	30	0.50	0.01	0.42**
	After	30	0.08	0.07	
6. Light sensitivity	Before	30	0.50	0.01	0.08**
	After	30	0.42	0.13	
7. Color	Before	30	0.51	0.01	0.00
	After	30	0.50	0.11	

Table 3: *T-test comparing the average silverfish from each simulation, before and after simulations.*

Looking at Table 4 where the results from *with* and *without* the human, and measured after simulation, are compared. The only difference between these simulations are observed in gene 6.

Gene	Human	N	Mean	Std.dev	Mean difference
1. Lifespan	No	15	0.73	0.09	-0.01
	Yes	15	0.74	0.10	
2. Starting energy	No	15	0.88	0.05	0.01
	Yes	15	0.87	0.04	
3. Perception	No	15	0.51	0.06	-0.01
	Yes	15	0.52	0.09	
4. Reproduction	No	15	0.13	0.13	-0.01
	Yes	15	0.14	0.09	
5. Rep. cycle	No	15	0.08	0.09	0.00
	Yes	15	0.08	0.04	
6. Light sensitivity	No	15	0.47	0.17	0.10*
	Yes	15	0.36	0.06	
7. Color	No	15	0.49	0.08	-0.02
	Yes	15	0.51	0.13	

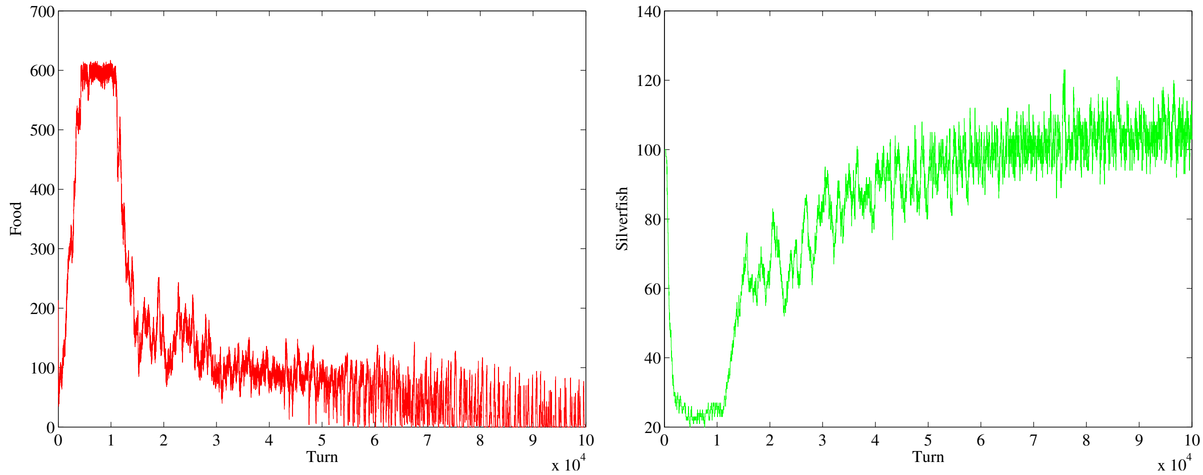
Table 4: *T-test comparing the average silverfish from 30 end of simulations with and without the human.*

5.2 Secondary results

The following results can be viewed as an indication if the simulation has been successful or not, i.e. that the relationship between food and silverfish seems reasonable but also that there existed an environmental pressure during the simulation - specifically the lack of food and the threat of the human.

5.2.1 Relation between food and silverfish

For a typical simulation when a human was *not* present, Figure 5a shows the amount of food available on the map in relation to elapsed time, the corresponding figures for the amount of silverfish in relation to time is presented in Figure 5b. The correlation between the amount of silverfish and food is as one could expect, strong and negative ($r = -0.92^{**}$ [4]). In other words, when the amount of silverfish increases, the amount of food decreases - and vice versa.

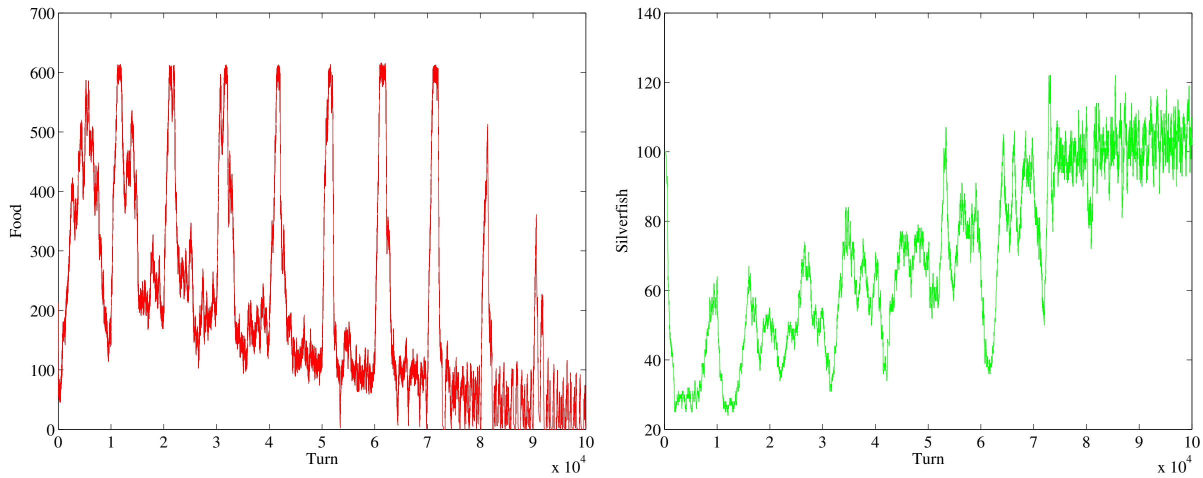


(a) *Amount of food during each turn* (b) *Amount of silverfish during each turn*

Figure 5: *Simulation without human*

Figure 6 shows the results corresponding to Figure 5 for a typical simulation *with* a human present. Again there is as one would expect a strong negative correlation between the amount of silverfish and food ($r = -0.80^{**}$). Note the spikes corresponding to the periods of time when the human was present, every 1000 turns. Also note how this is reflected in the production of food and how these figures differ from Figure 5.

[4] r is the Pearson's correlation coefficient.



(a) *Amount of food during each turn* (b) *Amount of silverfish during each turn*

Figure 6: *Simulation with human*

5.2.2 Causes of death

Table 5 shows the average numbers for the causes of death. The averages are from each simulation. The amount of silverfish dying from starvation is considerably larger than from age or killed by the human. However the standard deviations are quite large indicating a large variation between simulations.

	Age		Starvation		Human	
	Mean	**Std.dev**	**Mean**	**Std.dev**	**Mean**	**Std.dev**
1. Total	957.83	312.37	6667.37	2675.63	183.10	212.87
2. Without human	1022.60	293.14	6560.47	2296.97	0	0
3. With human	893.07	327.42	6774.27	3086.88	366.20	148.42

Table 5: *The average cause of death for all simulations* (1)*, only simulations without the human* (2) *and with the human* (3)*. Each mean is the average number of dead silverfish per simulation.*

6 Discussion

When comparing the results with the hypotheses stated beforehand, these can be categorized into conclusions regarding how the genes have changed after evolution, if evolution has indeed taken place and if so, if it has been repeated when running multiple simulations, and lastly regarding the success of simulation.

6.1 Genetic evolution

Looking at the figures for the whole population in Table 1 and Table 2, they're more or less consistent with the stated hypotheses. The means for the genes concerning lifespan and starting energy level increased, while the genes concerning energy level needed for reproducing and the reproductive cycle decreased as predicted (category 1). In other words it was more favorable for survival and reproduction to live longer, be born with a higher energy level, reaching an age of sexual maturity early and be able to reproduce more often. Also as expected, color didn't seem to be an important evolutionary factor in any simulation (category 3). However, the gene concerning perception didn't increase as expected (category 1). It means that the ability to detect food and other silverfish was not as important as predicted for survival and reproduction.

It's more plausible that the gene concerning perception had no effect, rather than that the mean in the tables is an optimal value for the gene, as both the mean and the standard deviation has barely changed after the simulation, but also because the mean and standard deviation is close to those of the control gene, color. A possible explanation for this is that the size of the map was too small in relation to the amount of food and silverfish, i.e. even a silverfish with limited perception could more or less easily come close enough to food and other silverfish simply by moving at random. This means that the range for which perception has an effect is lower than we anticipated, i.e. if there is only an effect of perception when it is below e.g. 0.30 then it would not matter if perception is at 0.50 or 0.90. A way to test this hypothesis would be to start simulations with a lower value for perception or perhaps rather decrease the maximum possible perception from 60 to e.g. 30.

As anticipated the gene concerning sensitivity to light barely changed when the human wasn't present, however it *decreased* when the human was present contrary to the hypothesis that it would *increase*. In other words it was favorable to be *less* sensitive to the light in relation to the starting values, i.e. not favorable to detect the human from a greater distance. As discussed later on under *Simulation*, it seems the impact of the human was indirect, leading to silverfish dying of *starvation* rather than killed by the human. This implies that a limited ability to detect the human was more favorable, which is also a plausible explanation for the result of evolution on this gene.

6.2 Replication of results

The most interesting figures are found in Table 3 and Table 4 as these are indicators of both the consistency of the results *between* simulations, but also better indicators if genes have evolved differently between conditions.

Looking at Table 3, genes belonging to category 1, with the exception of the gene concerning perception, a difference before and after simulations was observed. This was also the case for the gene belonging to category 2 concerning sensitivity to light. More importantly though, looking at Table 4 in relation to this gene, there was a difference between simulations when the human was present and not present. While for the rest of the genes no such differences could be observed. For the gene belonging to category 3, concerning color, and the gene concerning perception, no differences could be observed in neither condition. For the gene belonging to category 3, this result was expected.

These results supports the theory that silverfish are indeed optimized to their living environment. Although perception was not affected, the main part of the genes belonging to category 1 show the anticipated differences supporting the theory that silverfish are changing due to adaptation. Further support for this is that the gene belonging to category 2 showed different results when simulating with and without the human threat, while no differences were observed for genes belonging to category 1 and 3. Lastly as the results from a number of simulations showed similar results one can conclude that indeed evolution was successfully simulated and silverfish were optimized to their living environment.

6.3 Success of simulation

The figures concerning the relationship between the amount of silverfish and the amount of food looks as one would expect with a strong negative correlation. Also as one would expect the most common cause of death was starvation in Table 5, it seems that a limitation of food also limited the silverfish population which implicates that limited food supply acted as an environmental pressure as anticipated.

However the figures regarding the simulations with a human from Table 5 are a bit surprising. Looking at Figure 6 there is a clear indication that the human *had* an effect on the population of silverfish, but the number of silverfish killed by the human is not even half of the amount of silverfish which died of old age. A possible explanation for this is that as avoiding the human was the most prioritized course of action for a silverfish, the risk for starvation would be higher, i.e. a silverfish detecting the human from a large distance will not eat nor reproduce. As the human was present for 1000 turns each time he or she appeared and the maximum possible energy level of a silverfish was 2500, a silverfish avoiding the human may well be at greater risk of starvation. It seems plausible in relation to both Figure 6 and Table 5 that the impact of the human is rather that silverfish starve to death, as a result from avoiding the human, than are killed by the human directly.

6.4 Odd results

In both Figure 6, Figure 5 and in all the simulations we have observed, the population size is sharply reduced at the beginning of the simulation up until 800-1200 turns into the simulation where it began to increase again. In approximately 1 of 5 simulations the population dies out completely. Shown in the figures, the population is reduced to about 20 individuals and this is more or less representative to the other simulations - however we have no statistics to show this. The decrease indicates that there is a relatively large proportion of combinations of starting values that are unfavorable for survival, a risk we tried to reduce by sampling the values from a normal distribution. On the other hand, it's hard to say what is a low figure and not. We should not start from the optimal values because we want to let evolution take its course and leave room for both increases and decreases. These results could however be an indication that the starting values could be chosen more wisely. Especially so when also considering the aforementioned results in relation to maximum values for the genes related to perception and sensitivity to light. It's possible that choosing starting values closer to the ones gotten in this study would give a more stable simulation and clearer results.

As seen in Table 5 the standard deviation for cause of death generally is high in relation to means. Meaning that in some simulations notably more or notably less silverfish died, comparing with the average. A possible explanation is that the silverfish surviving the first 1000 turns can be more or less fit to the environment and this has a seemingly large effect on the continued simulation. This might further be an indication that starting values should be chosen more wisely.

There was a perhaps odd detail in Figure 3, a relatively large number of silverfish with a value of 1.00. This might be explained by the fact that there was an upper limit for the value, if there hadn't been one, these values would most probably be distributed roughly around the normal distribution.

6.5 Possible improvements

Lower maximum values for perception and sensitivity to light should be chosen to see if the theories of why we got the results we did hold. Otherwise we see mainly two areas of possible improvements for a more stable simulation and more conclusive results, these are the choosing of starting values and how the simulation is terminated.

As previously mentioned there are indications that starting values could have been chosen more wisely. It would be interesting to start the simulation with similar normal distributions to the ones used but with means corresponding to the optimal values gotten in this study and see how and to what extent these values change. Some of the genes should reach it's maximum or minimum range at some point in time as the optimal value should be maximum or minimum, e.g. lifespan and reproductive cycle respectively. This can be further investigated by choosing some other termination criteria.

For this study the simulations lasted 100 000 turns as this was decided before start of simulation. It would be interesting to see how genes evolve if a termination criteria linked to the changes of the genes were set instead of a time limit. In other words, terminating the simulation once the genes more or less stops evolving. This could possibly also be a better indication of optimal fitness compared to the results gotten in this study.

6.6 Self-criticism

One could ask if we merely produce the results we want or that our hypotheses influence our results? This question can be answered both yes and no. Looking at what evolution is it consists of a deterministic environment (at least from a macro perspective to avoid discussions involving quantum physics etc.) that shapes the genes of organisms by natural selection. As we are creators of the environment in this simulation we *do* shape the genes of the silverfish, but once the environment is created and the simulation is started we are no longer in control so the influence is indirect. Stochastic events determines how genes are changing and by natural selection the environment shapes the genes of the silverfish.

The fact that we create the environment means it would be ignorant to *not* have hypotheses, pretending that the environment is not created after a preplanned model. It would be equally ignorant to ignore that our hypotheses can affect the results of this study, the best we can do is state these hypotheses in beforehand and hope that our awareness of them might reduce this effect. In that way there is at least a possibility to assess to what extent our hypotheses may have affected the results. Also note that we have not calibrated the genes to achieve the results we expect, calibration is done relative to achieving a stable simulation and mainly concerns reproduction of food, map size, population size and the human.

6.7 Summary

Although all hypotheses were not correct, the results show that it's possible to simulate evolution by modelling it as an evolutionary algorithm. The fact that a number of simulations showed similar results and that simulation was conducted using different environmental conditions supports the idea that the silverfish were optimized to its' environment through evolution or natural selection.

These results support the idea of using evolutionary algorithms for optimization, but it is also of support to the notion of complex organisms arising by relatively simple mechanisms as stated by the theory of evolution.

References

[1] T. Back, U. Hammel, and H.-P. Schwefel. Evolutionary computation: Comments on the history and current state. *Evolutionary computation, IEEE Transactions on*, 1(1):3–17, 1997.

[2] R. Brandon. Natural selection. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Center for the Study of Language and Information, Standford University, spring 2014 edition, 2014. http://plato.stanford.edu/archives/spr2014/entries/natural-selection/.

[3] A. E. Eiben and M. Schoenauer. Evolutionary computing. *Information Processing Letters*, 82(1):1–6, 2002.

[4] R. Lewontin. The genotype/phenotype distinction. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Center for the Study of Language and Information, Standford University, summer 2011 edition, 2011. http://plato.stanford.edu/archives/sum2011/entries/genotype-phenotype/.

[5] J. Misra. Distributed discrete-event simulation. *ACM Computing Surveys (CSUR)*, 18(1):39–42, 1986.

[6] H. Perros. Computer simulation techniques: the definitive introduction. *Computer Science Department, North Carolina State University, Raleigh, North Carolina, USA*, pages 1–3, 71–72, 2003. http://www4.ncsu.edu/~hp/simulation.pdf.

[7] D. Tiechroew, J. F. Lubin, and T. D. Truitt. Discussion of computer simulation techniques and comparison of languages. *Simulation*, 9(4):182–183, 1967.